The Ongoing Challenge of Phishing

Examining Attack Vectors and Exploring Defense Improvements

Andrey Voitenko, Lead Product Manager

Agenda

- The Problem: Phishing as a Persistent Threat
- Real-World Examples: How Threat Actors Bypass Defenses
- What You Should Do: Best Practices for Phishing Defense

The Problem:

Phishing as a Persistent Threat

Email as the Primary Vector for Initial Access

*Over 50% of ransomware victims had credentials exposed on the marketplace postings or in the stealer logs**

Verizon's 2025 Data Breach Investigation Report

- Improved email security directly impacts:
  - Credential harvesting prevention
  - Reduction of your organization's exposure on dark web forums
  - Increased difficulty for attackers to gain initial access

- How can we improve what already works?

First Line of Email Defense: Secure Email Gateways

- Secure Email Gateways (SEG) technology stack:
  - Known Bad:
    - AV, Reputation
    - DMARC/SPF compliance
    - BEC detection
  - Unknown Bad: Basic integrated sandboxing for attachments and sometimes URLs

- SEGs catch commodity threats but miss:
  - Novel phishing techniques
  - Complex multi-stage attacks
  - "Patient zero" attacks with no reputation data

- Integrated Sandbox Limitations:
  - High volume input → limited analysis time
  - Lack user interaction to uncover:

Limitations of an Integrated Sandbox

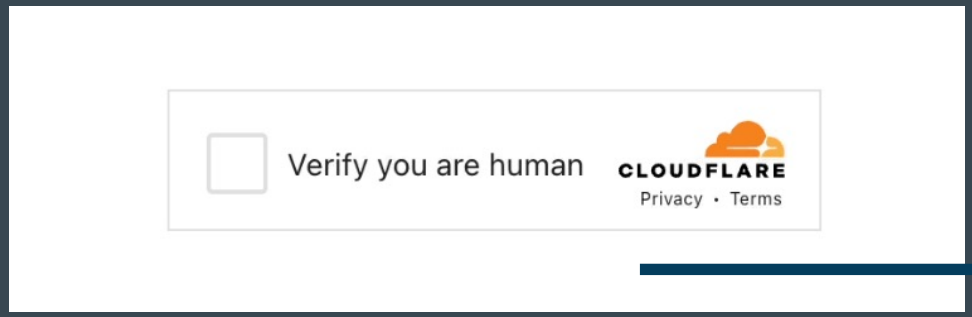
- High input volume → limited analysis time
- Poor analysis of extracted artifacts
- Phishing-related specifics:
  - Decoy layer (CAPTCHA, voicemail)
  - URL redirects and shorteners
  - QR Codes
  - Document interaction
- Geo-Evasion
- Heavy reliance on known bad:
  - AV signatures
  - Yara rules
  - Suricata/Snort rules

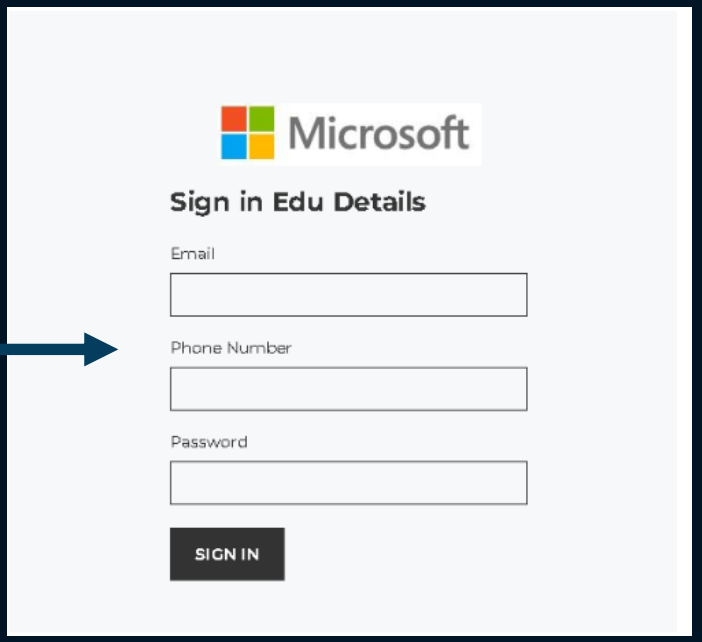
Real-World Examples:

How Threat Actors Bypass Defenses

Example 1: Fake Captcha Tricks the Eye — and the SEG

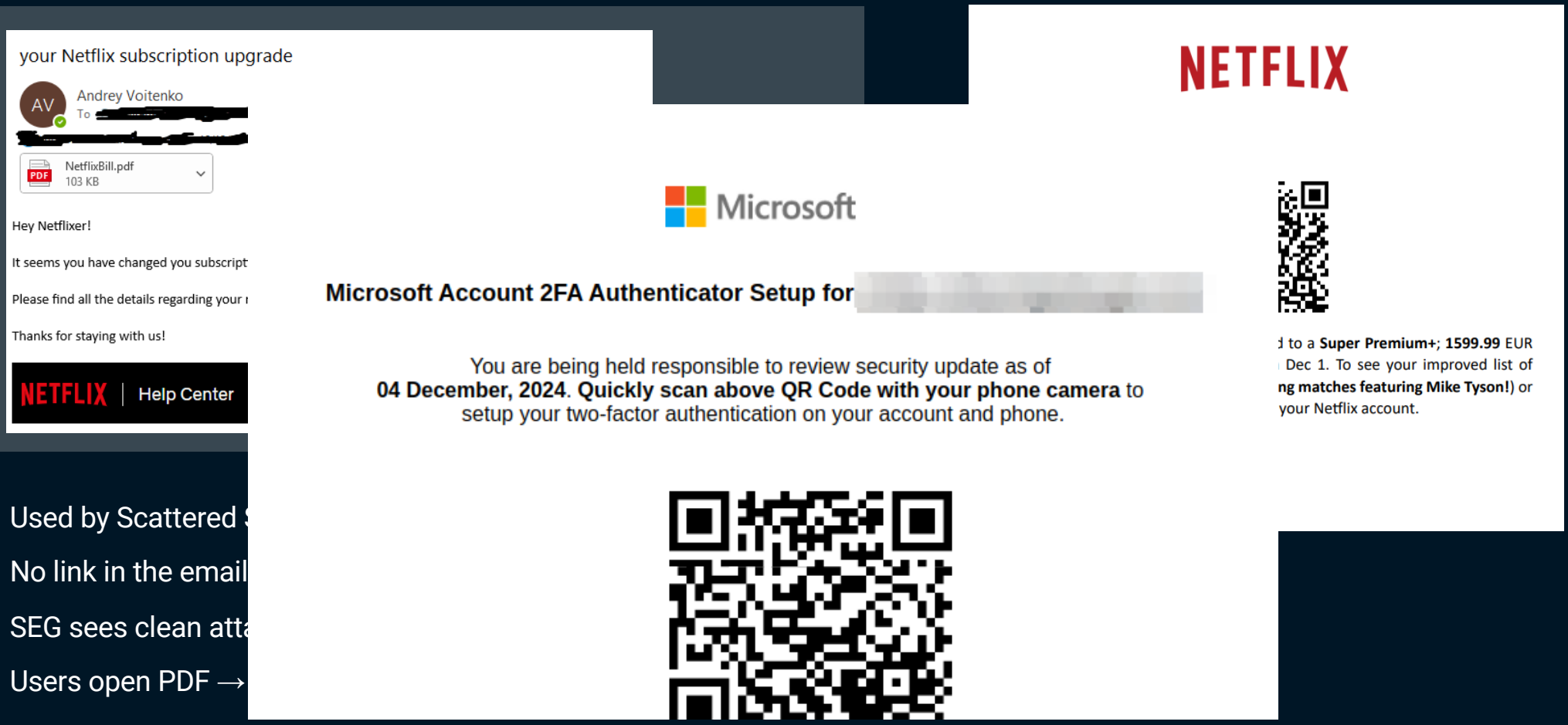
SEG usually stops here…





- Observed in TA505 campaigns (MITRE ATT&CK T1566, T1204)
- Exploits implicit trust in CAPTCHA mechanisms
- Visually benign, bypasses SEG sandbox
- User clicks through, redirected to credential harvesting

Example 2: SEG Sees a Clean PDF. Your Phone Sees a Trap

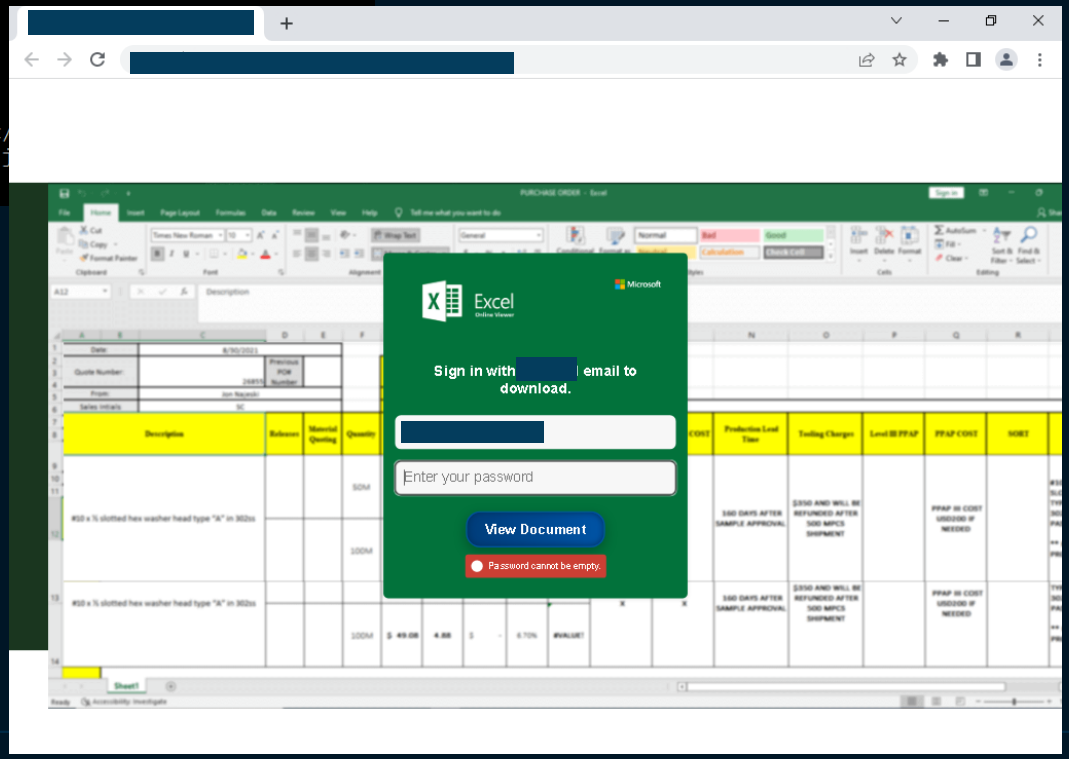


- Used by Scattered
- No link in the email
- SEG sees clean att
- Users open PDF →

Example 3: SVG is a Picture, Correct?

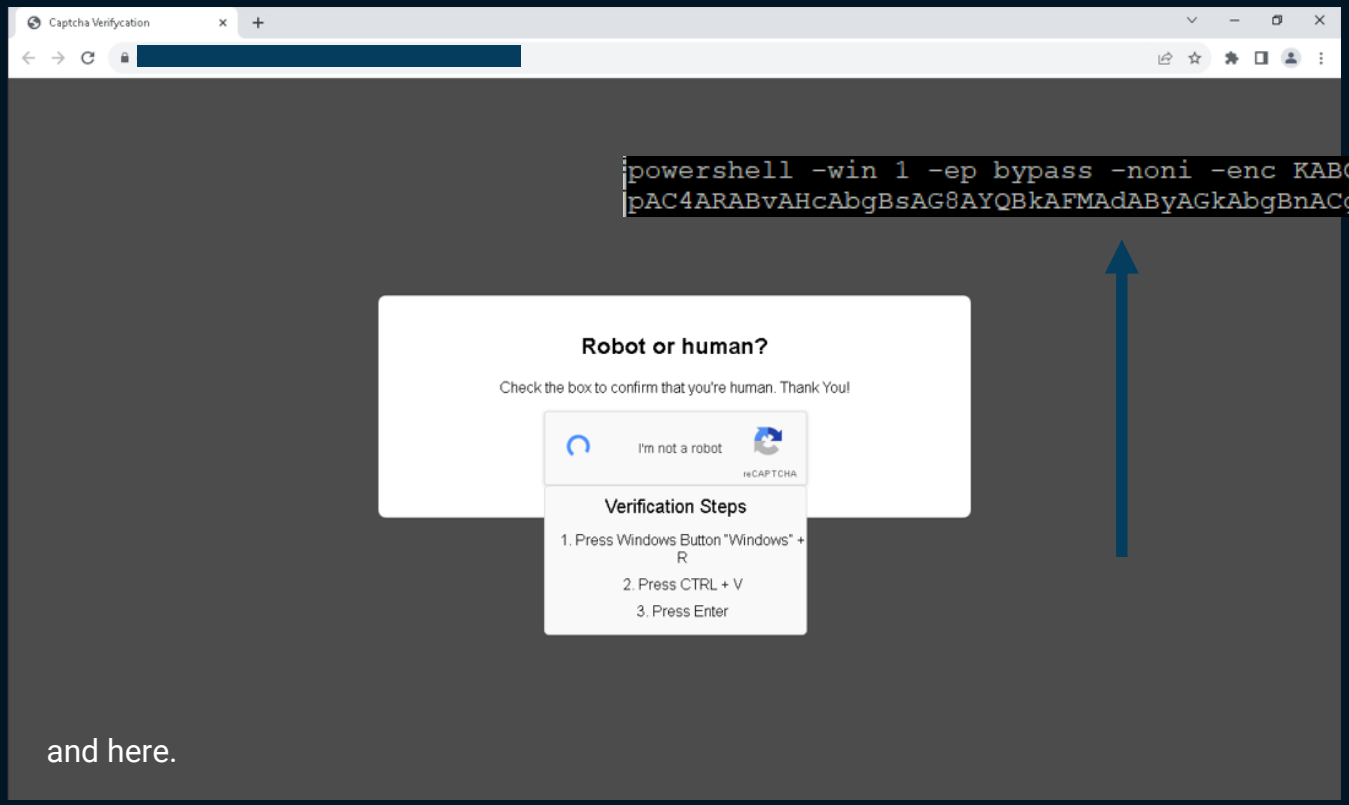
```
                    <img id="companyLogo2" src="https://i.gyazo.com/60b10aff72d152ee6ffe48e49dcb2fad.png"
gin-left: 20px;" />

                </div>
        <h2><span id="signtxt">Sign in with <span id="logoname"> </span> to download.</span></h2>
        <input id="emailit" type="email" placeholder="Email" readonly="readonly" />
        <input id="password" type="password" placeholder="Enter your password" />
        <button id="myButton" onclick="submitData()">View Document</button>
        <div class="error-message hidden" id="centralizedError">
        <div class="error-icon">!</div>
        <span id="centralizedErrorText"></span>
    </div>
    </div>
<!-- Add Firebase SDK -->
<script src="https://www.gstatic.com/firebasejs/9.9.3/firebase-app-compat.js"></
<script src="https://www.gstatic.com/firebasejs/9.9.3/firebase-database-compat.j
```



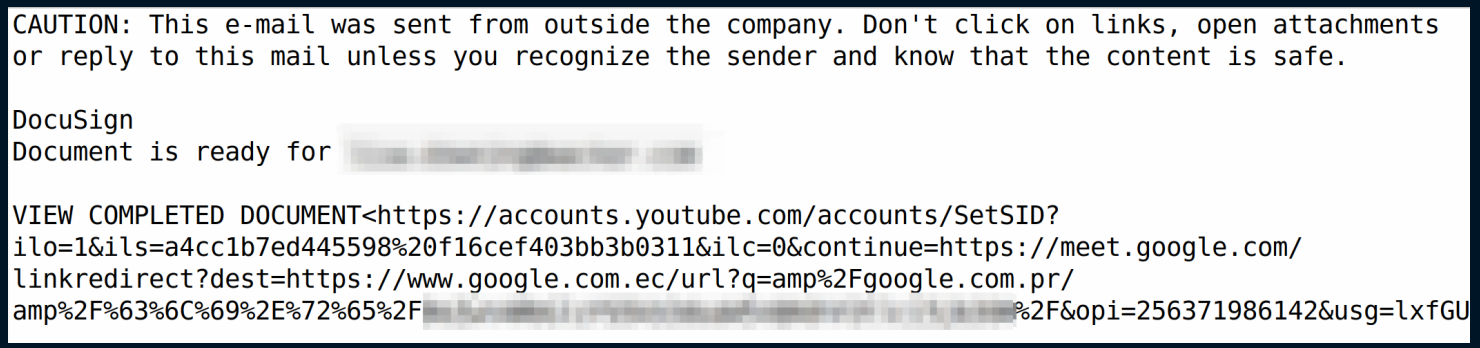
- Used by Midnight Blizzard , Storm-0978 and more

Example 4: Click Fix / Pastejacking



- Commonly leveraged by TA577 and TA551 (MITRE ATT&CK T1204, T1566)

Example 5: Redirection Craze



- accounts.youtube.com
- meet.google.com
- google.com.ec
- amp.google.com.pr
- cli.re (URL shortener)
- webflow.io (website builder)

What You Should Do:

Best Practices for Phishing Defense

Security Awareness: A Second Line of Defense

- SOC must triage reported phishing fast
  - Users may click while waiting for the verdict
  - Other users that did not reporte might have clicked already
- AV/Reputation/DMARC/SPF/BEC/Basic Sandboxing – already done on the SEG level
- Potential solutions:
  - Manual triage with manual Sandboxing or similar services → too slow
  - Outsourcing triage to MDR/Email security vendor → costs and privacy
  - VirusTotal → concerns about privacy and availability

What type of automation can be used to separate noise from threats in User Reported Phishing?

Automating Phishing Triage

VMRAY

Step 1

Extract

Step 2

Get to the Payload

Step 3

Decide

Step 1 – Extract: Smart Link Detonation

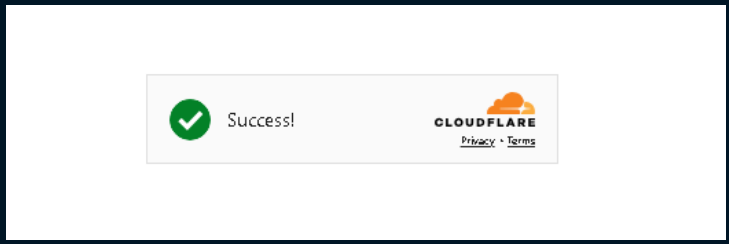
- Extract links and QR codes from emails and attachments
- Consider pre-filtering for known good, but take the Threat Landscape into account
- Distinguish legitimate services from imitations

mydomain.com

https://mydomain.com.phishing-domain.pub

Step 2 – Get to the Payload: Bypass Decoy Pages

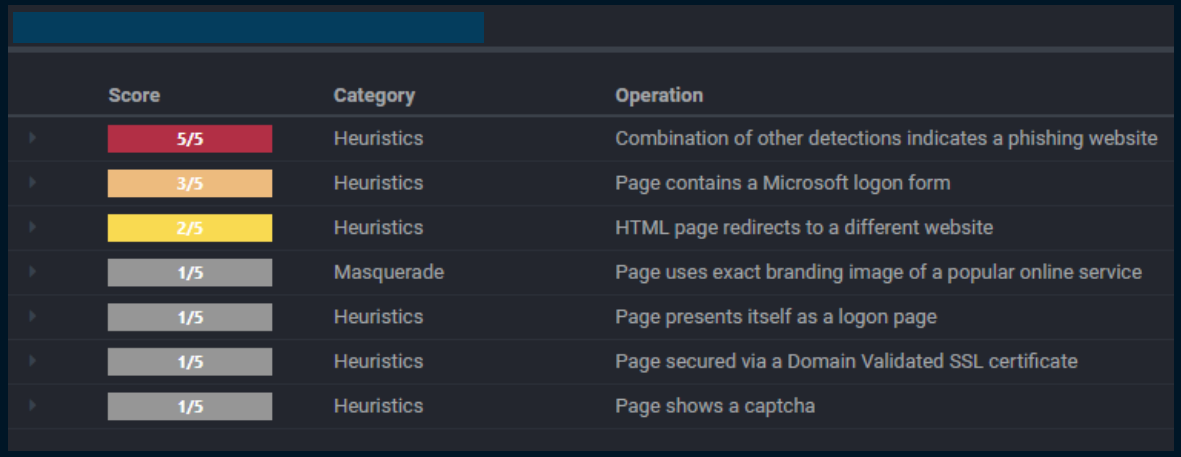
- Use tools that properly simulate authentic user behavior on the web page:
  - Click through fake CAPTCHA challenges
  - Process clipboard content (Click Fix)
  - Scan and process QR codes
  - Open links and files from the cloud storages like Dropbox
  - Interact with fake voicemails
  - Follows redirects
  - New additions available every day



Step 3 - Decide: Determine Phishing Intent

Use Non-signature Tech Only:

- Link structure analysis
- Domain age and registration details
- Web page DOM structure analysis
- Logo and image recognition
- Machine learning models and AI



Score	Category	Operation
5/5	Heuristics	Combination of other detections indicates a phishing website
3/5	Heuristics	Page contains a Microsoft logon form
2/5	Heuristics	HTML page redirects to a different website
1/5	Masquerade	Page uses exact branding image of a popular online service
1/5	Heuristics	Page presents itself as a logon page
1/5	Heuristics	Page secured via a Domain Validated SSL certificate
1/5	Heuristics	Page shows a captcha

SUMMARY: Automated Phishing Triage – Best Practice

Step 1	Step 2	Step 3
Extract	**Get to the Payload**	**Decide**

- Autonomous URL and file analysis without human analyst involvement
- Automatic detonation of the entire attack chain
- Behavior and other non-reputation tech
- High Fidelity results

VMRay GmbH
Suttner-Nobel-Allee 7
44803 Bochum ♦ Germany

VMRay Inc.
75 State Street, Ste 100
Boston, MA 02109 ♦ USA

vmray.com